AFRL-RI-RS-TR-2014-068

SCREEN FINGERPRINTS AS A NOVEL MODALITY FOR ACTIVE AUTHENTICATION

UNIVERSITY OF MARYLAND

MARCH 2014

FINAL TECHNICAL REPORT

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

STINFO COPY

AIR FORCE RESEARCH LABORATORY
INFORMATION DIRECTORATE

■ **AIR FORCE MATERIEL COMMAND** ■ **UNITED STATES AIR FORCE** ■ **ROME, NY 13441**

NOTICE AND SIGNATURE PAGE

Using Government drawings, specifications, or other data included in this document for any purpose other than Government procurement does not in any way obligate the U.S. Government. The fact that the Government formulated or supplied the drawings, specifications, or other data does not license the holder or any other person or corporation; or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

This report is the result of contracted fundamental research deemed exempt from public affairs security and policy review in accordance with SAF/AQR memorandum dated 10 Dec 08 and AFRL/CA policy clarification memorandum dated 16 Jan 09. This report is available to the general public, including foreign nationals. Copies may be obtained from the Defense Technical Information Center (DTIC) (http://www.dtic.mil).

AFRL-RI-RS-TR-2014-068 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

FOR THE DIRECTOR:

/ S /
ANNA WEEKS
Work Unit Manager

/ S /
WARREN H. DEBANY, JR.
Technical Advisor, Information
Exploitation & Operations Division
Information Directorate

This report is published in the interest of scientific and technical information exchange, and its publication does not constitute the Government's approval or disapproval of its ideas or findings.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
MARCH 2014	FINAL TECHNICAL REPORT	MAY 2012 – OCT 2013

4. TITLE AND SUBTITLE

SCREEN FINGERPRINTS AS A NOVEL MODALITY FOR ACTIVE AUTHENTICATION

5a. CONTRACT NUMBER
FA8750-12-2-0199

5b. GRANT NUMBER
N/A

5c. PROGRAM ELEMENT NUMBER
61722F

6. AUTHOR(S)

Rama Chellappa

5d. PROJECT NUMBER
ATAU

5e. TASK NUMBER
UM

5f. WORK UNIT NUMBER
AR

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Maryland
Room No 4411 A. V. Williams Building
Department of Electrical and Computer Engineering
College Park, MD 20742-3275

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Air Force Research Laboratory/RIGA
525 Brooks Road
Rome NY 13441-4505

10. SPONSOR/MONITOR'S ACRONYM(S)
AFRL/RI

11. SPONSOR/MONITOR'S REPORT NUMBER
AFRL-RI-RS-TR-2014-068

12. DISTRIBUTION AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited. This report is the result of contracted fundamental research deemed exempt from public affairs security and policy review in accordance with SAF/AQR memorandum dated 10 Dec 08 and AFRL/CA policy clarification memorandum dated 16 Jan 09.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

A screen fingerprint is proposed as a new biometric modality for active authentication. Such a fingerprint is acquired by taking a screen recording of the computer being used and extracting discriminative visual feature from the recording.

15. SUBJECT TERMS
Active authentication, screen fingerprints, biometrics

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			**ANNA WEEKS**
U	U	U	UU	16	19b. TELEPHONE NUMBER (*Include area code*) **N/A**

TABLE OF CONTENTS

Section		Page
1.0	**Summary**	**1**
2.0	**Introduction**	**1**
3.0	**Methods, Assumptions, and Procedures**	**3**
3.1	Advantages over Language-based Techniques	3
3.2	Advantages over Motor-based Techniques	4
3.3	Advantages over Application-based Techniques	4
4.0	**Results and Discussion**	**5**
4.1	Dataset	5
4.2	Features: Average Histogram of Oriented Optical Flows	6
4.3	Interaction Classification	7
4.4	Experimental Evaluations	7
4.5	Identity Verification	7
4.6	Experimental Evaluations	8
5.0	**Conclusions**	**10**
6.0	**References**	**11**

1.0 SUMMARY

We investigate if screen-based recordings of computer interactions can be used for accurate user authentication. A dataset of screen recordings of some PC inter- actions (MouseMoving, Typing, Scrolling, Other) of 21 users was collected and we ran a set of experiments to help our investigation. We extract low-dimensional feature vectors based on histogram of optical flows from each screen recording. The first set of experiments investigated if these low-dimensional features can be used to recognize the type of interaction taking place in a particular recording and we found that a linear SVM could succeed in achieving this with an accuracy of 91%. The second set of experiments explored if classifiers trained on different types of recordings can be used to continuously verify a user identity. The results indicated that SVMs trained on Scrolling recordings can achieve moderately low FAR and FRR error rates. These results indicate that further research in using screen-based recordings for active authentication can lead to a more reliable cyber biometric.

2.0 INTRODUCTION

Biometrics deals with the problem of identifying individuals based on physiological or behavioral characteristics. Since many physical characteristics, such as face, iris, etc., and behavioral characteristics, such as voice, expression, keystroke, etc., are unique to an individual, biometric analysis offers a reliable and natural solution to the problem of identity verification. It has been shown that physiological biometrics techniques have been more successful for the problem of identity verification than behavioral characteristics. This is due in part to the fact that physiological features remain stable for long periods of time. On the other hand, behavioral characteristics are greatly influenced by one's mood, stress or illness. This makes it somewhat instable for identity verification.

The current standard method for validating a user's identity for authentication of a computer device requires humans to do something that is inherently difficult: create, remember, and manage long, complex passwords. Furthermore, as long as the session remains active, typical systems incorporate no mechanisms to verify that the user originally authenticated is the user still in control of the computer. Thus, unauthorized individuals may improperly obtain access to the computer if a password is compromised or if a user does not exercise adequate vigilance after initially authenticating on a device.

To deal with this problem, various cyber biometrics have been proposed in the literature. These methods capture the cognitive fingerprints of users. The proposed theory is that how individuals formulate their thoughts and actions are reflected through their behavior, and this behavior in turn can be captured as metrics in how the individual performs tasks using the computer. The most notable examples are keyboard dynamics [6] and mouse dynamics [9]. Some other examples of the computational behavior metrics of the cognitive fingerprint include eye tracking, how

the user selects information, how the user searches for information, etc. We have proposed a novel way of validating the identity of the person using a computer that focuses on the unique aspects of the individual through the use of screen fingerprint. Screen fingerprint is the new cyber biometric modality that we have proposed to measure and analyze active authentication. The screen finger- print is acquired by taking a screen recording of the computer being used by the operator and by extracting discriminative visual features from these recordings.

The screen fingerprint of an operator captures enough of the unique human qualities to be usable as a biometric for authentication. The qualities captured include cognitive abilities, motor limitations, subjective preferences, and work patterns. For example, how well the operator sees is a cognitive ability that can be captured visually by the size of the text shown on the screen. How fast the operator drags a window is a motor limitation that can be captured visually by the amount of motion detected on the screen. How organized the operator arranges multiple windows is a subject preference that can be captured visually by the layout of salient edges identified on the screen. What suite of applications the operator uses is a work pattern that can be captured visually by the distribution of application- specific visual features recognized on the screen.

The proposed technology exploits the synergy between recent advances in pixel-level screen analysis [4], [13], [14] and vision-based biometrics. Vision- based biometrics such as face and iris recognition has become more reliable. Yet, its dependence on hardware sensors often limits its applicability. On the other hand, pixel-level screen analysis has received a lot of attention in human computer interaction in the past two years. One attractive advantage of pixel-level screen analysis is its wide applicability, since the screen buffer can be accessed on all platforms at the software level. However, pixel-level screen analysis has not been used as a modality for biometrics. This is the first attempt to combine vision- based biometrics and pixel-level screen analysis in a complementary manner for the purpose of active authentication.

In order to study the effectiveness of screen fingerprints, we put together a dataset in which a significant number of screen recordings of different individuals are collected under various conditions. We describe this dataset and present some results using two state-of-the-art classifiers. Based on our experiments using this dataset, we found that indeed screen fingerprints can capture enough of the unique human qualities to be usable as a biometric for active authentication. We demonstrate that in addition to applying a good classification algorithm, finding features that are robust to variations present in screen recordings are very important for authentication.

3.0 METHODS, ASSUMPTIONS, AND PROCEDURES

Over the past few years, computer vision techniques have been successfully applied to the analysis of the graphical user interfaces shown in a screen recording to support a wide range of applications including automation [12], search [14], software testing [4], and tutorial [13]. Some applications perform batch analysis after screen recordings are acquired, such as searching online documentation about the interfaces in a screen recording [14], [13]. Some applications operate in real-time while screen recordings are made. For example, Yeh et al. has developed the Sikuli visual automation tool [12] that can observe a screen recording in real-time, identify an interface component by appearance, and send automation command (e.g., click) to that component. This tool has had a significant impact on software engineering in that it is currently used by dozens of companies to automate GUI testing. Active authentication based on screen fingerprints is a novel application of screen recording analysis that has never been attempted before.

A typical scenario of active authentication using screen fingerprints proceeds as follows. First, an operator of the computer logs on using an initial authentication mechanism such as entering a password. While the operator is using the workstation, the screen of the computer is being observed. Screen recording are taken within short observation windows. Each time a screen recording is taken, the recording (a video) is visually analyzed to extract a screen fingerprint aimed to identify the person who is using the computer. This observed screen fingerprint is compared to the reference screen fingerprint previously measured and stored for the authorized operator. If a match is established between the observed and reference fingerprints, the operator is actively authenticated. Now suppose the operator steps away and leaves the workstation unattended. An adversary may gain physical access to the computer. While the adversary is using the computer, a screen recording is taken to extract a screen fingerprint. However, the observed screen fingerprint no longer matches the reference screen fingerprint. As a result the workstation may lock itself up to prevent further unauthorized use by the adversary.

Screen fingerprints offer several advantages over other potential modalities for active authentication as described below:

3.1 Advantages over Language-based Techniques

Language-based techniques such as those based on computational linguistic and structural semantic analysis seek to authenticate computer operators based on verbal cues such as the words and phrases an operator uses in digital communication (e.g., emails, memos). These stylometry techniques [2], [3] do not work well in situations when operators' primary responsibilities do not involve personal communication (e.g., data entry) or when operators mainly use mouse or touch-based interfaces (e.g., Photoshop). Our proposed modality can deal with these situations because it relies on

visual cues that are always observable on a computer screen regardless of the types of applications operators use.

3.2 Advantages over Motor-based Techniques

Motor-based techniques seek to authenticate computer operators based on kinetic cues such as how fast an operator types [11] or moves a mouse pointer [10]. These techniques cannot support operators who use voice or touch as the primary input modality. Our proposed modality can authenticate operators who do not use a mouse or keyboard because it does not depend on specific input devices.

3.3 Advantages over Application-based Techniques

Application-based techniques seek to authentication computer operators based on usage cues such as which applications or features an operator is using. However, these techniques are difficult to scale because each application must be specifically instrumented in order to track its usage. Often such deep instrumentation requires access to the application's source code or special application programming interface (API). Comprehensive coverage is hard to attain because some proprietary or legacy applications do not provide source code or API for instrumentation. Our proposed modality is able to provide wide coverage over most applications without deep instrumentation. As long as an application is visible on a computer screen, it can be captured in a screen recording. Some of the distinctive visual properties of an application can be extracted to be part of an operator's screen fingerprint.

4.0 RESULTS AND DISCUSSION

4.1 Dataset

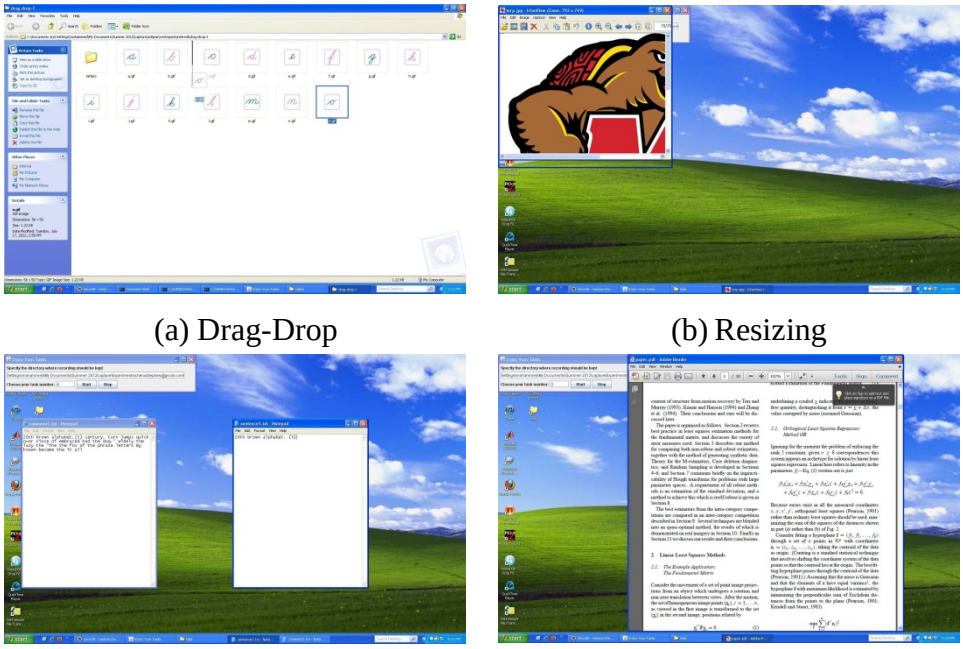
We collected a dataset of screen recordings of PC interactions of 21 users. Each user was asked to perform 4 types of tasks: dragging icons, typing, scrolling and resizing. The user was directed to repeat the different tasks 5 times but in a permuted order. We developed a java program to guide the user through the data collection process. The program recorded screen at a rate of 12 frames per second. We found that rate did not affect the system responsiveness and was good enough to capture the visual dynamics of the different types of interactions. The 4 tasks collected are described as follows:

Drag-Drop: The user is asked to drag a set of files into a certain directory, one file at a time. This is an instance of the MouseMoving interaction type.

Resizing: The user is asked to resize an image window so that it takes (roughly) the left half of the screen. This is also an instance of the MouseMoving interaction type.

Typing: The user is asked to type specified paragraphs into a typing window. Clearly, this is an instance of the Typing interaction type.

Scrolling: The user is asked to scroll through a document and count the number of times a certain letter appears in the section titles. The type of interaction here is Scrolling.



(a) Drag-Drop (b) Resizing

(c) Typing (d) Scrolling

Figure 1: Screenshots of different tasks.

Figure 1 shows a sample screenshot from each task. Before starting the actual task, there may be some frames where the user is switching context to start doing the task. Similarly, there can be some frames after finishing the actual task where the user is terminating the current task. Accordingly, we manually specified for each recorded video the start and end times within which the core interaction takes place. This divides each recording into a sub-video of the core interaction type (Typing, MouseMoving or Scrolling) and up to 2 sub-videos of Other interaction. We end up with 1243 instances of the 4 different types of interactions (Other, Typing, MouseMoving and Scrolling). The average recording length in seconds is 2 for Other, 43 for typing, 12 for MouseMoving, and 83 for Scrolling. The typical data collection session for a particular user lasted between 20-25 minutes.

4.2 Features: Average Histogram of Oriented Optical Flows

After taking a screen recording, the next step is to extract a feature vector that deally should distinguish different kinds of interactions (e.g. Scrolling vs Mouse- Moving) and discriminate different users (legitimate vs illegitimate) while being inexpensive to compute. One of the most popular techniques of measuring the change in visual appear- ance between two consecutive frames f t and f t+1 is the optical flow wt. It can be thought of as a velocity field over the image that describes the visual motion at every pixel [7]. We believe that features based on optical flow can well discriminate among different interaction kinds. Indeed, Scrolling is characterized by vertical (and/or horizontal) visual motion whereas MouseMoving is characterized by continuous motion in all directions rather than just vertical or horizontal. Optical flow can also discriminate different users as it is sensitive to their different interaction rates.

We use the following procedure to extract a feature vector x from a screen recording r = {f0, f1, ..., fT −1}. First, we downscale each frame ft to 160 x 100 resolution and use the downscaled frames to calculate the optical flows
{w0, w1, ..., wT −2} using the implementation in [8]. We then calculate a Histogram of oriented Optical Flow (HOF), denoted by ht, from each wt using the implementation in [5]. We restrict the number of orientation bins of ht to 100 but we neither normalize the histograms nor subdivide the images into grids as in [5]. Finally, we compute x as the arithmetic mean of all the HOF vectors:

$$x = \frac{1}{T-1}\sum_{t=0}^{T-2} h_t$$

For all the experiments reported in this report, we use the 100D AHOF vectors extracted from the screen recordings.

4.3 Interaction Classification

We need to train a classifier that can be used to automatically determine the type of interaction occurring in a screen recording. Each recording in the dataset is labelled not only by its user ID but also by the type of interaction. This allows us to use supervised machine learning techniques to do the training. The two techniques we tried are Support Vector Machine (SVM) [1] and AdaBoost [15]. For SVM, we used the soft-margin, linear kernel version and we handled multiclasses using the One-Versus-All (OVA) strategy where we trained a separate SVM to classify each interaction type against all the other interaction types. For AdaBoost, we used the single-node decision tree as the weak classifier and handled multiclasses by implementing the simple variation of AdaBoost called SAMME described in [15].

4.4 Experimental Evaluations

We used the interactions of 14 randomly selected users to train both SVM and AdaBoost. The parameter tuning was performed on the interactions of 2 other users and the testing was done on the interactions of the remaining 5 users. The classification accuracy was as high as 91% for SVM versus 80% for AdaBoost. The confusion matrices of both techniques are also shown in Table 1 and 2, re- spectively. In both matrices, each row indicates that most of the matching errors result from confusing the row's interaction type as Other (e.g. 6% of the Mouse- Moving instances in Table 1 are misclassified only as Other). This is explained by the fact that there are many more instances in the Other class than all other inter- action classes. This type of error (i.e. non-Other instances being misclassified as Other) is more acceptable in practice than the other types of errors (e.g. instances misclassified as non-Other) because other types of errors will result in invoking a verification classifier on a recording of an incompatible type (e.g. running the Scrolling-based verification classifier on an Other recording).

4.5 Identity Verification

We experimentally investigated whether all interaction types are equally powerful in verifying user identity. We measure the detection error metrics commonly used in the evaluation of biometrics. These are defined below:

- FAR: False Acceptance Rate is the fraction of illegitimate samples (i.e. negatives) that are incorrectly accepted (i.e. classified as positives). The complementary fraction of FAR is the True Rejection Rate (TRR = 1 - FAR). Maximizing TRR is equivalent to minimizing FAR.

- FRR: False Rejection Rate is the fraction of legitimate samples (i.e. positives) that are incorrectly rejected (i.e. classified as negatives). The complementary fraction of FRR is the True Acceptance Rate (TAR = 1 - FRR). Maximizing TAR is equivalent to minimizing FRR.

When a biometric is evaluated, it typically assigns a real score to each instance in the test set. These scores are then mapped into decisions (accept/reject) based on a real threshold θ. Different thresholds can lead to different detection error rates (FAR, FRR) (or equivalently accuracy rates (TRR, TAR)). We select the threshold that achieves the best tradeoff between FAR and FRR. To do this, we define a new performance metric that we call F1-DET (short for F1 of Detection Error Tradeoff). F1-DET is simply the harmonic mean of TRR and TAR. The idea of F1-DET is analogous in concept to the traditional F1 score and how it is used to conservatively evaluate different (precision, recall) pairs. In other words, F1-DET is high only when both TRR and TAR are high (or equivalently FAR and FRR are both low). If either TRR or TAR or both are low, the corresponding F1-DET will also be low.

4.6 Experimental Evaluations

We tried different configurations of K-Nearest Neighbours (KNN) and soft-margin SVM as biometric classifiers. For each user u and interaction c, we train a biometric to verify the legitimate instances in S(u, c) against the illegitimate instances in S(c) − S(u, c). We do this by running a K-fold cross validation process with K = |S(u, c)|. The data set S(c) is divided into K parts where each part contains one legitimate (positive) sample from S(c, u) and 1/K of the illegitimate (negative) samples in S(c) − S(c, u). In the ith fold, we train on all parts except the ith, which is used for testing. After completing all folds, we evaluate FAR, FRR and F1-DET at all possible threshold values and set F1D(u, c) to the highest F1-DET score. In addition, we set FAR(u, c) and FRR(u, c) to the pair corresponding to F1D(u, c).

Table 3 shows for each classifier and interaction class c the scores F1D(c), FAR(c), and FRR(c) where F1D(c) is the average of F1D(u, c) taken over all users u. FAR(c) and FRR(c) are defined in a similar fashion. It is easy to see that Scrolling leads to the best detection accuracy compared to other classes of interaction. This is true for all classifiers although Scrolling performance is best with SVM. Typing has the lowest detection accuracy (i.e. F1D) compared to other types of interaction. Figure 2 shows for each interaction c the box-plot of the F1D(u, c) scores achieved by the classifier that was found to give the highest aver- age F1D score for interaction c in Table 1. The Figure leads to observations similar to those derived from Table 1. In addition, it shows that the worst F1D scores achieved by the linear SVM for Scrolling tend to be better than those achieved by the RBF-Kernel SVM although Table 3 indicated that both have the same average F1D score for Scrolling.

Table 1: Performance of user verification using different classifiers and different interactions

Classifier	Other			Typing		
	F1-DET	FAR	FRR	F1-DET	FAR	FRR
KNN (K=3, Weighted)	69.85%	25.80%	33.29%	53.76%	46.90%	39.05%
KNN (K=3, Unweighted)	**74.38%**	15.10%	27.77%	54.94%	47.00%	22.86%
KNN (K=1)	68.84%	18.91%	28.09%	57.48%	53.33%	5.71%
Linear SVM	70.61%	21.85%	33.82%	59.96%	37.57%	36.19%
RBF-Kernel SVM	62.42%	39.57%	33.10%	**63.43%**	31.05%	37.14%

Classifier	MouseMoving			Scrolling		
	F1-DET	FAR	FRR	F1-DET	FAR	FRR
KNN (K=3, Weighted)	63.71%	33.87%	36.88%	77.93%	21.90%	18.10%
KNN (K=3, Unweighted)	**69.85%**	24.38%	27.67%	76.98%	20.43%	13.33%
KNN (K=1)	63.10%	26.81%	27.62%	72.04%	31.76%	9.52%
Linear SVM	68.76%	30.29%	29.89%	**82.75%**	20.67%	12.38%
RBF-Kernel SVM	64.52%	40.00%	27.83%	**82.75%**	18.81%	14.29%

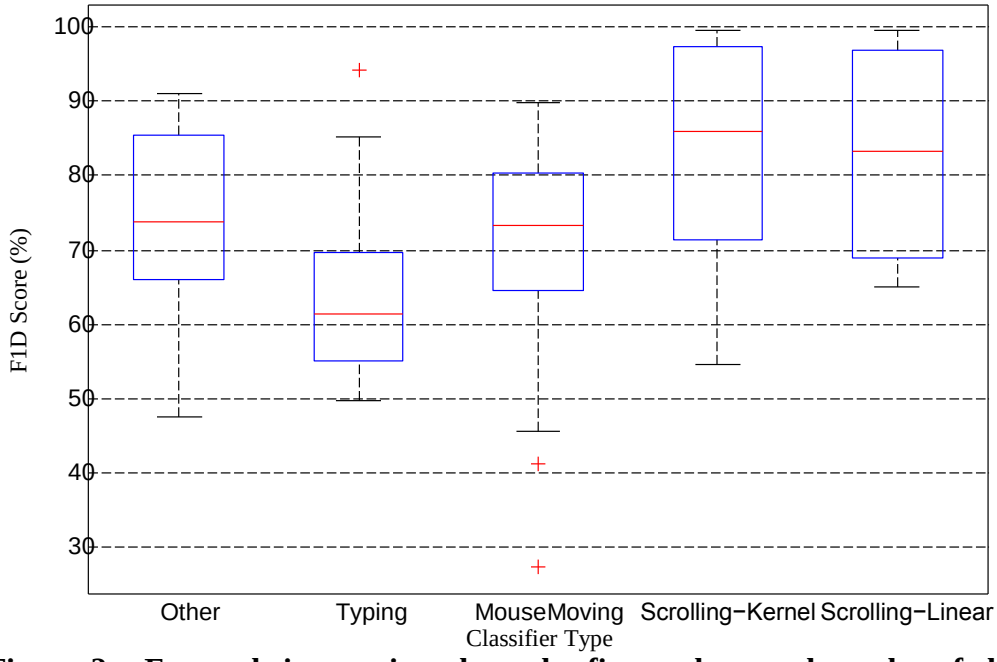


Figure 2: For each interaction class, the figure shows a box plot of the F1D scores of all uses achieved by the classifier that was found best for that interaction class. That is, it shows the distribution of the F1D scores of KNN (K=3, Unweighted) for Other, RBF-Kernel SVM for Typing, KNN (K=3, Unweighted) for Mouse- Moving, RBF-Kernel SVM and Linear SVM for Scrolling.

5.0 CONCLUSIONS

Our results indicate (1) that screen recordings of Scrolling lead to moderately low detection error rates (FAR = 20.67% and FRR = 12.38%) and (2) that not all classes of interactions are equally reliable. Based on our results, a screen- based biometric should be activated only during Scrolling interactions as other interactions are less reliable. Although the verification performance obtained may not be as high as some of the other longer-established modalities such as mouse dynamics, screen output can enhance the security of a multi-modal system in case there is little of the data that other modalities monitor. It is also worth noting that the performance of the other modalities have been the target of research for much longer time (33 years for keystroke dynamics [6] and 9 years for mouse dynamics [9]) and the performance of screen fingerprints can be further improved beyond the reported results by investigating richer features and other classifiers.

6.0 REFERENCES

[1] Christopher M Bishop. *Pattern recognition and machine learning*, Volume 1. Springer, 2006.

[2] K Calix, M Connors, D Levy, H Manzar, G MCabe, and S Westcott. Sty- lometry for e-mail author identification and authentication. *Proc. Student- Faculty CSIS Research Day, Pace University*, 2008.

[3] Omar Canales, Vinnie Monaco, Thomas Murphy, Edyta Zych, John Stewart, Charles Tappert, Alex Castro, Ola Sotoye, Linda Torres, and Greg Truley. A stylometry system for authenticating students taking online tests. *Proc. Student-Faculty Research Day, CSIS, Pace University*, 2011.

[4] T.-H. Chang, T. Yeh, and R. Miller. GUI testing using computer vision. In *Proceedings of the Conference on Human Factors in Computing System*, may 2010.

[5] R. Chaudhry, A. Ravichandran, G. Hager, and R. Vidal. Histograms of ori- ented optical flow and binet-cauchy kernels on nonlinear dynamical systems for the recognition of human actions. In *Proc. IEEE Conf. on Comput. Vision*

[6] R Stockton Gaines, William Lisowski, S James Press, and Norman Shapiro. Authentication by keystroke timing: Some preliminary results. Technical report, DTIC Document, 1980.

[7] B.K.P. Horn and B.G. Schunck. Determining optical flow. *Artificial intelli- gence*, 17(1):185–203, 1981.

[8] C. Liu et al. *Beyond pixels: exploring new representations and applications for motion analysis*. PhD thesis, Massachusetts Institute of Technology, 2009.

[9] M. Stamp S. Hashia, C. Pollett. On using mouse movements as a biometric. In *Proc. Int'l Conf. Computer Science and its Applications*, pages 1–8, 2004.

[10] Chao Shen, Zhongmin Cai, Roy A Maxion, Guang Xiang, and Xiaohong Guan. Comparing classification algorithm for mouse dynamics based user identification. In *Biometrics: Theory, Applications and Systems (BTAS), 2012 IEEE Fifth International Conference on*, pages 61–66. IEEE, 2012.

[11] Terence Sim and Rajkumar Janakiraman. Are digraphs good for free-text keystroke dynamics? In *Proc. IEEE Conf. on Comput. Vision and Pattern Recognition (CVPR)*, pages 1–6, 2007.

[12] T. Yeh, T.-H. Chang, and R.C. Miller. Sikuli: using GUI screenshots for search and automation. In *Proceedings of the 22nd annual ACM symposium on User interface software and technology*, pages 183–192, 2009.

[13] T. Yeh, T.-H. Chang, B. Xie, G. Walsh, K. Wongsuphasawat, I. Watkins, M. Huang, L.S. Davis, and B. Bederson. Creating contextual help for GUIs using screenshots. In *Proceedings of the 24th ACM Symposium on User Interface Software and Technology*, may 2011.

[14] T. Yeh, B. White, J. San Pedro, B. Katz, and L.S. Davis. A case for query by image and text content: searching computer help using screenshots and keywords. In *Proceedings of the 20th International Conference on World Wide Web*, may 2011.

[15] J. Zhu, S. Rosset, H. Zou, and T. Hastie. Multi-class adaboost. *Ann Arbor*, 1001(48109):1612, 2006.